

LA-UR-11-11458

Approved for public release; distribution is unlimited.

Title:	Evaluation of the Assessing Institutional Digital Assets (AIDA) Toolkit
Author(s):	Miller, Laniece E. Blake, Miriam E.
Intended for:	Science and Technology Libraries



Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By acceptance of this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Evaluation of the Assessing Institutional Digital Assets (AIDA) Toolkit

LANIECE MILLER*

MIRIAM BLAKE

MELANIE SORSBY

Los Alamos National Laboratory Research Library, Los Alamos, NM

Libraries are becoming more involved in the complex task of managing the digital assets created by members of their institutions. Many tools have been created to help librarians understand and solve the problems associated with this task.

One of these tools is the Joint Information Systems Committee's Assessing Institutional Digital Assets toolkit, which is designed to help institutions assess their current readiness and ability to manage digital assets. This article provides a review and evaluation of the AIDA toolkit as used by Los Alamos National Laboratory Research Library.

Keywords: Digital Asset Management Tools, AIDA Toolkit, Digital Curation, Data Management Planning

Running Head: Evaluation of the AIDA Toolkit

*Laniece Miller is a Graduate Research Assistant at the Los Alamos National Laboratory Research Library. PO. Box 1663, MS-P362, Los Alamos, NM, 87545.

Email: lemiller@lanl.gov

Miriam Blake is the Group Leader / Director of the Los Alamos National Laboratory Research Library.

Melanie Sorsby is a Librarian at the Los Alamos National Laboratory Research Library

Acknowledgements: We would like to thank Frances Knudson and Amber Wu for their input as we worked through the AIDA toolkit. Also Linn Collins, Matthew Hopkins, James Powell and Michael Wenman helped with editing of the paper. The project is funded by the Department of Energy.

Digital Asset Management

There has been much discussion both within and outside of the library community related to effective long term management of digital objects important to institutions. Records managers know that access to digital records is crucial for effective business operations (Heslop 2002), but digital asset management extends beyond just records and record management. Research libraries are beginning to address issues related to the digital asset management of scholarly material. This scholarly material may take the form of formal documents, images, video, software, or data files. Libraries are focused more specifically on the digital asset management of scholarly material through the development of processes intended to add value and maintain digital information so that it can remain available to researchers into the future. These actions are often framed as “digital curation.” (Beagrie, 2006.)

The role of libraries in society is changing partially due to the fact that they are shifting work from management of traditional, physical materials to meeting the challenge of digital asset management (Angevaere 2009). Currently, libraries may manage digital content in three primary ways; providing access to metadata and electronic full-text for publisher or vendor content, managing digitized local collections, and managing institutional, scholarly digital assets. For publisher data, libraries mainly host metadata and negotiate contracts, leaving preservation of the electronic full-text to the publisher. Digitized local special collections tend to be historical collections of materials which the library has digitized in order to provide better access. Scholarly digital assets created by the institution, including research data, are the

newest challenge and often considered the most challenging task because of the complexity of the objects which need to be organized and stored in a sustainable manner (Angevaere 2009).

Managing institutional digital assets is important for several reasons, including the fragility of digital objects, the exponential growth of digital output, and the potential to support collaboration, especially in the sciences. Unlike print, benign neglect is not an option for digital assets because physical media decay, digital files can become corrupted, and hardware and software becomes obsolete (Carpenter 2005). A 2000 study conducted by the University of California at Berkeley estimated 93% of intellectual output is produced digitally (Kenney 2003). Data is a uniquely complex type of digital asset, ranging from small datasets (e.g. spreadsheets) to huge data collected by sensors and machines. A dataset may include multiple forms of digital objects linked together. Many of the “big science” projects produce data on the order of petabytes per year, causing unique data management concerns (Gray 2011). Discussion of a possible fourth scientific paradigm, data driven science, had indicated a requirement for easy access to existing data (Nelson, M.L. 2009). Databases such as the Protein Data Bank allow for the easy sharing of expensive or irreplaceable data (Beagrie 2006).

Datasets in particular have come to the forefront of discussion because some funding sources and publishers have begun to require data management plans as part of the grant proposal process. The US National Science Foundation (NSF) and National Institute of Health (NIH) both have data management requirements in the proposal process (National Science Foundation 2011, National Institutes of Health 2003). Some journals, such as *Science*, *Nature*, and journals published by the American Geophysical Union (AGU) require access to research data (American

Association for the Advancement of Science 2011, Nature Publishing Group 2011, American Geophysical Union 2011). In 2011, the Department of Energy (DOE) updated DOE Order 241.1B to announce of all useful and available scientific and technical information (STI) to the Office of Scientific and Technical Information (OSTI), not just the traditional publications (US Department of Energy, 2010). Datasets are specifically mentioned.

With requirements coming from several different sources, some libraries are helping to meet these new information management needs. Librarians are accustomed to organizing information, especially scholarly output, so digital asset management, including data, is a natural extension (MacDonald 2008). Some argue that data repositories should be organized along disciplinary lines, not institutional. There is increasing acknowledgment that research libraries can play vital roles including support for faculty by organizing, providing access, and potentially housing discipline based repositories or smaller interdisciplinary datasets, and providing metadata and access services (Messerschmitt 2003). Also libraries could play roles in managing digital assets which are interdisciplinary and are outside traditional disciplinary lines. For librarians trained in traditional areas of information science, such as metadata and reference services, attempting to engage in e-science means learning more about the data life cycle, the fragility of digital objects, archiving, and digital preservation (Gold 2007). Many resources regarding this subject for librarians are becoming available both in the peer-reviewed literature and as helpful tools such as websites and planning toolkits.

Awareness of new requirements regarding managing digital data prompted staff at the Los Alamos National Laboratory (LANL) Research Library to learn more about how to help

researchers manage digital scientific assets pertaining to their research. Research Library staff began the learning process by completing a self-assessment using the Assessing Institutional Digital Assets (AIDA) toolkit to determine the library's readiness to manage digital assets at the Laboratory. Digital asset management is a fairly recent Laboratory-wide undertaking, and there is no centralized digital asset management system or policy established. The self assessment was a way to gauge what issues would need to be addressed to move forward with a strategy. From the outset, the LANL team recognized that the AIDA toolkit was written from an academic library perspective and that DOE libraries differ from academic research libraries in areas such as size, organizational structure, institutional mission, and funding sources. Therefore, this paper attempts to give a general overview and critique of the AIDA toolkit, specifically from the perspective of the LANL Research Library.

Planning Tools

Planning is a vital step towards digital asset management and preservation for future users (del Pozo, 2010). To aid in planning, multiple organizations have created toolkits evaluating various aspects of digital asset management. These toolkits generally fall into one of two categories: data profiling tools and audit or assessment tools. The data profiling tools provide guidance on communicating with researchers to understand their data management needs. The audit or assessment toolkits are a top down review of current digital asset management. Working together these tools can aid libraries in effective digital asset management planning.

Data profile tools such as Purdue's Data Curation Profiles and the Data Asset Framework (DAF) provide a way to formally determine how researchers creating digital content view the issue of

digital management and what needs they have surrounding the issue. Many researchers are concerned about uncontrolled use (or misuse) of their data, causing much to be hidden and potentially lost. To help overcome these fears and integrate with their work, researchers must be involved in creating any successful solutions. This lesson has been learned by the creators of many institutional repositories around the country that are now sitting almost empty due to lack of researcher support and other barriers (Nelson, B. 2009). Since data curation is defined as the management of data through the lifecycle of interest, data must include not just numerical data but the associated digital output created during the research process (Carlson 2010). This digital output could be as software, scientific workflows, experimental settings, etc. created during the research process. Purdue's Data Curation Profiles toolkit was designed to help information professionals record the researchers' point of view regarding their digital management needs (Witt 2009). The DAF is a survey toolkit primarily written for information professionals who understand the data lifecycle (Data Asset Framework: Implementation Guide 2009).

The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit may be the audit tool with the most name recognition. According to the first iteration of the toolkit, "The DRAMBORA toolkit represents the latest development in an ongoing international effort to conceive criteria, means and methodologies for audit and certification of digital repositories" (DCC 2007). DRAMBORA is an audit designed as an evaluation tool for existing repositories, not as a planning guide. The Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) is a self-audit tool created by RLG and the National Archives and Records Administration (NARA). Like DAMBORA, TRAC's goal is to develop criteria to certify existing repositories as

trustworthy for long term curation and management of digital assets (Dale 2007). As audits, these toolkits require documentation and can be overwhelming for a library in the planning stages, but can still be useful in setting tactical goals.

The AIDA toolkit is designed to assess an institution's current readiness to manage digital assets. AIDA is not an audit but a self-assessment; its target is not the evaluation of an established repository but an assessment of the culture of the institution or organization (Pincent 2009). The AIDA project team, funded by the Joint Information Systems Committee (JISC) in the UK, is now working on the Integrated Data Management Planning Toolkit & Support (Pincent September 2010). Two complete versions and a revision of the first AIDA toolkit are available on the AIDA project site (Pincent AIDA: A JISC toolkit). Because LANL does not have a laboratory wide repository, the library used the AIDA toolkit as a first step. The LANL library used the revision of the first version, released February 2009.

AIDA – Evaluation of the Toolkit

The AIDA toolkit was developed by JISC and the University of London Computer Center (ulcc) in the United Kingdom. The goal of the AIDA toolkit is to aid institutions and departments in determining how prepared they are to manage their important digital assets for the long term.

The 2009 revision of the toolkit focuses on two different organizational levels: the entire institution and the smaller sub-unit such as an individual division, department, or group. The LANL Research Library used the toolkit differently, focusing on two aspects of the scientific digital assets managed by the library itself. One aspect was digital content the library manages to provide researcher access to scholarly content, such as the locally loaded commercial article

database, the library catalog and local publication databases. The second aspect was digital objects created for internal use by the library staff, such as procedures and workflows. The management of digital assets at the institutional level was not reviewed as part of this initial process.

The AIDA toolkit is divided into three sections, termed legs, based on Cornell's maturity model (Pincent 2009, Kenney 2005). The organizational leg includes the non-technical areas: policy, metadata, legal concerns, asset sharing, and audit trails. The technology leg covers issues pertaining to the technology components of the infrastructure, information integrity, metadata, and disaster planning. The resources leg covers the resources dedicated to digital asset management including funding, staffing, and business planning (Pincent 2009). The toolkit covers many of the issues which a library needs to consider in the planning stages of digital asset management. Each of the three legs is split into several elements, totaling 31, each of which are scored as one of Cornell's five stages: acknowledge, act, consolidate, institutionalize, and externalize (Pincent 2009). The goal of the team performing the self-assessment is to read the element and level descriptions and determine where the institution and/or department falls. Each element has a number, brief name, and descriptions of each stage for that element.

The AIDA toolkit was specifically designed for use by the UK academic institutional community (Pincent *AIDA: A JISC Project*) which is a subset of research libraries. The toolkit may not be an exact fit for other institutions so the team completing the toolkit may have to determine if some elements do not apply or if other concepts should be considered. For example, the LANL Research Library team found the idea of needing guidance, or even a policy, for appraisal to

determine which scientific digital assets should be retained was not addressed in the toolkit. On the other hand, the Laboratory as a whole regulates computer security very strictly, so the element in the toolkit addressing information security was unnecessary because of the rigorous standards already in place at LANL.

Each element gives descriptive text for the stages, leaving the team to interpret what is being evaluated. With so much freedom to interpret many of the elements, the team can easily become consumed with how to apply the toolkit. As an example, the description of the organizational leg element “Digital management and sharing” appears to jump between a discussion about the strategy for sharing files and a discussion of the system where they are actually shared. Another difficulty was that both the institutional and departmental levels were listed together on the same sheet, unlike the newer 2010 version. The large amount of textual information on the page forces the group doing the assessment to consciously evaluate what applies to their task. The team completing the assessment must understand that intensive thought and discussion are necessary to successfully complete the toolkit.

Elements in different legs can be extremely similar to each other, such as a metadata element which occurs in both the organizational and technology legs. The organizational element attempts to evaluate if there is an appropriate schema for assets being managed. The technology leg discusses the automation of the metadata collection process. Unfortunately these differences are often murky and unclear, thus forcing the group to spend time defining the differences between elements rather than assessing the stage of readiness to provide a robust digital asset management capability at the institution.

Despite these difficulties, the LANL Library found AIDA helpful in determining current strengths and weaknesses pertaining to their ability to manage digital assets. Not having to supply supporting documentation, such as policy documents that confirm internal decisions, simplifies the toolkit and makes it accessible for a library in the planning process. In contrast, the other assessment tools assume that an existing repository is being audited and require supporting documentation. AIDA can be completed by a small group in a few sessions around a conference table with nothing more than the toolkit and a way to take notes. The LANL Research Library took one session for each leg, of about an hour to an hour and a half each.

The AIDA toolkit covers a broad spectrum of issues within digital management without assumptions, and is useful in uncovering current strengths and weaknesses in an organization. AIDA forces the assessors to think about managing digital assets beyond just the apparent technical aspects. Once strengths and weaknesses are known, a plan can be formed or improvements can be made to existing strategies. At LANL, the tool has proved most valuable for its ability to help start conversations regarding managing digital assets with other stakeholders. Since a library must form partnerships both within the institution and outside it, discussions with other stakeholders about specific requirements or toolkit elements can help build necessary relationships.

Completing the Toolkit and Moving Forward

To be successful in completing the toolkit, there are several key points to keep in mind while working through the assessment. One size does not fit all, meaning that not everything in the toolkit applies to every institution or can be easily interpreted at first glance. The team must

approach it with an open mind and be willing to make adjustments as needed. A team must have a clear driver for completing the toolkit and know what aspect of the institution is being evaluated to shape how the elements are interpreted. Keeping the goals in mind while completing the toolkit should relieve some of the stress caused by the ambiguity in the toolkit language. Because the issue is complex, the toolkit will take serious thought and discussion. The team members must expect the exercise of completing the toolkit to be challenging and also to challenge their current perceptions of digital asset management at their institution.

After completing the toolkit, a list of strengths and weaknesses can fairly easily be compiled. With such a list, a team working towards effective management of digital assets can begin to plan internally and initiate conversations with other stakeholders beyond the library and across the institution on the complex issues surrounding digital asset management. Some aspects may not be covered sufficiently in the toolkit but still need to be considered in planning. The exact process in moving forward will differ for every institution, based on the current digital asset management environment, and how their end goals are defined. The AIDA toolkit is designed to assist institutions in determining their readiness to manage digital assets. Working with tools like the AIDA toolkit can play a significant role in educating librarians and others in the wide array of issues that are critical to understanding and moving forward in the digital age.

References

- American Association for the Advancement of Science. 2011. "General information for authors." Accessed August 24, 2011.
http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail
- American Geophysical Union. 2011. "Policy on referencing data in and archiving data for AGU publications." Accessed August 24, 2011.
http://www.agu.org/pubs/authors/policies/data_policy.shtml
- Angevaere, I. 2009. Taking care of digital collections and data: "Curation" and organizational choices for research libraries. *LIBER Quarterly*. 19 (1): 1-12.
- Beagrie, N., 2006. Digital curation for science, digital libraries, and individuals. *The International Journal of Digital Curation*, 1 (1): 3-16.
- Carlson, J. 2010. *The Data Curation Profiles Toolkit: User Guide*. Purdue University Libraries.
<http://www.datacurationprofiles.org>
- Carpenter, L. 2005. Supporting digital preservation and asset management in institutions. *Ariadne*. 43 (April).
- Dale, R.L., et.al. 2007. *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Chicago, Illinois: CRL and OCLC.
- Data Asset Framework: Implementation Guide*. Digital Curation Centre. 2009.

del Pozo, N., Long, A.S., Pearson, D. 2010. "Land of the lost": a discussion of what can be preserved through digital preservation. *Library Hi Tech*. 28 (2): 290-300.

Digital Curation Centre, DigitalPreservationEurope. 2007. *DCC and DPE Digital Repository Audit Method Based on Risk Assessment, b.1.0*. retrieved December 15, 2010.

<http://www.repositoryaudit.eu/download>

Gold, A. 2007. Cyberinfrastructure, data, and libraries, part 1: A cyberinfrastructure primer for librarians. *D-Lib Magazine*. 13 (9/10). doi:10.1045/september20september-gold-pt1.

Gray, N., Woan, G. 2001. Digital preservation and astronomy: Lessons for funders and the funded. *Proceedings of Astronomical Data Analysis Software and Systems XX*. 442 : 13-16.
arXiv:1103:2318v1.

Heslop, H., Davis, S., Wilson, A. 2002. *An Approach to the Preservation of Digital Records*.
National Archives of Australia. 43 (April).

Kenney, A.R., McGovern, N.Y. 2003. The five organizational stages of digital preservation.
Digital Libraries: A Vision for the 21st Century: A Festschrift in Honor of Wendy Lougee of on the Occasion of her Departure from the University of Michigan. (Ann Arbor, MI: MPublishing, 2003) Accessed July, 19, 2011.

<http://hdl.handle.net/2027/spo.bbv9812.0001.001>

Kenney, A.R., McGovern, N.Y. 2005. "Three Legs, Two Documents, and Five Stages: Developing Your Organizational Response to Digital Preservation Requirements" at the Library and Archives Canada.

MacDonald, St., Uribe, L.M. 2008. Libraries in the converging worlds of open data, e-research, and web 2.0. *Online*. 32 (2): 36-40.

National Institutes of Health, 2003. *Final NIH Statement on Sharing Research Data*. Accessed August 23, 2011. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>. NOT-OD-03-032.

National Science Foundation, 2011. *Award and Administration Guide*. Accessed August 23, 2011. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4. Chapter VI.D.4.

Nature Publishing Group. 2011. "Availability of data and materials". Accessed August 24, 2011. <http://www.nature.com/authors/policies/availability.html>

Nelson, B. 2009. Empty Archives. *Nature*, 461 (10): 160-163.

Nelson, M.L. 2009. Data-driven Science: A new paradigm?. *EDUCAUSE Review*, 44 (4): 6-7.

Pinsent, E. 2009. *The AIDA Self-Assessment toolkit: Mark II, February 2009*. London: University of London Computer Centre.

Pinsent, E. September 29, 2010. "Adapted AIDA for IDMP." *AIDA: A JISC Project* (blog), <http://aida.jiscinvolve.org/wp/2010/09/>

Pinsent, E. *AIDA: A JISC Project*. Accessed December 16, 2010. <http://aida.jiscinvolve.org/wp/toolkit/>

US Department of Energy, 2010. *Scientific and Technical Information Management*. DOE O
241.1B

Witt, M., Carlson, J., Brandt, D.S., Cragin. M. 2009. Constructing data curation profiles. *The
International Journal of Digital Curation*. 3 (4): 93-103.